# Project 2

Daniel Fredin, Junhan Li, & Eric Chen

## Problem 1

## Problem 3

### a)

#### ii) 10-fold Cross-Validation

#### Model Accuracy

```r
# Predict outcome using model from training data based on testing data
predictions <- predict(cvmodel, newdata=test)

# Create confusion matrix to assess model fit/performance on test data
con_matx <- confusionMatrix(data=predictions, test$COMMIT)
con_matx
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction no yes
##        no  13   4
##        yes  2  11
##
##                Accuracy : 0.8
##                  95% CI : (0.6143, 0.9229)
##     No Information Rate : 0.5
##     P-Value [Acc > NIR] : 0.0007155
##
##                   Kappa : 0.6
##
##  Mcnemar's Test P-Value : 0.6830914
##
##             Sensitivity : 0.8667
##             Specificity : 0.7333
##          Pos Pred Value : 0.7647
##          Neg Pred Value : 0.8462
##              Prevalence : 0.5000
##          Detection Rate : 0.4333
##    Detection Prevalence : 0.5667
##       Balanced Accuracy : 0.8000
```

```
##
##        'Positive' Class : no
##
```

## b)

## c)

```
compute_weibull_stats <- function(shape, scale) {
  # Compute the mean
  mean_value <- scale * gamma(1 + (1/shape))


  # # Compute the median
  # median_value <- scale * qweibull(0.5, shape, scale)


  # Compute the median
  median_value <- scale * (log(2)^(1/shape))


  # Compute the mode
  if (shape > 1)
    mode_value <- scale * ((shape - 1) / shape)^(1/shape)
  else
    mode_value <- 0

  # Compute the variance
  variance_value <- scale^2 * (gamma(1 + (2/shape)) - (gamma(1 + (1/shape)))^2)

  # Return the computed statistics as a named list
  return(list(mean = mean_value,
              median = median_value,
              mode = mode_value,
              variance = variance_value))
}

# lambda = scale
# K = shape

# Example usage
scale_param <- 6
shape_param <- 1

stats <- compute_weibull_stats(shape_param, scale_param)

# Accessing the computed statistics
mean_value <- stats$mean
median_value <- stats$median
mode_value <- stats$mode
variance_value <- stats$variance
```

```r
# Printing the computed statistics
cat("Mean:", mean_value, "\n")
```

```
## Mean: 6
```

```r
cat("Median:", median_value, "\n")
```

```
## Median: 4.158883
```

```r
cat("Mode:", mode_value, "\n")
```

```
## Mode: 0
```

```r
cat("Variance:", variance_value, "\n")
```

```
## Variance: 36
```

### Maximum likelihood estimator (MLE) of the parameters from the Weibull distribution

Maximum Likelihood Estimator (MLE) of $\sigma$ from the normal distribution assuming $\mu$ is constant.

#### Step 1: Probability Density Function for a normal distribution.

$$f(x; \lambda, k) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

#### Step 2: The likelihood function (L).

$$\begin{aligned} L_{\hat{x}}(\lambda, k) &= \prod_{i=1}^{n} \frac{k}{\lambda} \left(\frac{x_i}{\lambda}\right)^{k-1} e^{-\left(\frac{x_i}{\lambda}\right)^k} \\ &= \left(\frac{k}{\lambda}\right)^n \left(\frac{1}{\lambda^{k-1}}\right)^n \left(\prod_{i=1}^{n} x_i^{k-1}\right) \left(e^{-\sum_{i=1}^{n}\left(\frac{x_i}{\lambda}\right)^k}\right) \\ &= \frac{k^n}{\lambda^{nk}} e^{-\sum_{i=1}^{n}\left(\frac{x_i}{\lambda}\right)^k} \prod_{i=1}^{n} x_i^{k-1} \end{aligned}$$

#### Step 3: The log-likelihood function (ln(L)).

$$\ln L_{\hat{x}}(\lambda, k) = n \ln(k) - nk \ln(\lambda) - \sum_{i=1}^{n} \left(\frac{x_i}{\lambda}\right)^k + (k-1) \sum_{i=1}^{n} \ln x_i$$

**Step 4: Derivative of the log-likelihood function (ln(L)).**

$$\begin{aligned}
\frac{d\ln L}{d\sigma^2} &= \frac{d}{d\sigma^2}\left(-\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right) \\
&= -\frac{n}{2\sigma^2} - \left[\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right]\frac{d}{d\sigma^2}\left(\frac{1}{\sigma^2}\right) \\
&= -\frac{n}{2\sigma^2} - \left[\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right]\left(-\frac{1}{(\sigma^2)^2}\right) \\
&= \frac{1}{2\sigma^2}\left[\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 - n\right]
\end{aligned}$$

**Step 5: Solving for the desired parameter.**

$$\begin{aligned}
\frac{d\ln L}{d\sigma^2} &= 0 \\
\frac{d\ln L}{d\sigma^2} &= \frac{1}{2\sigma^2}\left[\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 - n\right] = 0 \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 - n = 0 \\
&= \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 = \sigma^2
\end{aligned}$$

Therefore the estimator $\hat{\sigma}$ is equal to the unadjusted sample variance.

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2}$$

Since $\hat{\mu} = \bar{x}$ then the estimator $\hat{\sigma}$ is equal to,

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$